LU02a - Normalverteilung

Ziel: Du kannst erklären, was eine Normalverteilung ist, die 68-95-99.7-Regel anwenden und Datenwerte mithilfe des **z-Scores standardisieren**.

Kurztheorie (Merksätze)

- **Normalverteilung:** Daten häufen sich um einen zentralen Wert und nehmen nach beiden Seiten gleichmässig ab → Glockenform.
- **Mittelwert = Median = Modus** (alle drei fallen zusammen).
- **68 %** aller Werte liegen innerhalb von **1 Standardabweichung** vom Mittelwert.
- **95 %** aller Werte liegen innerhalb von **2 Standardabweichungen** vom Mittelwert.
- **99.7 %** aller Werte liegen innerhalb von **3 Standardabweichungen** vom Mittelwert.
- Der **z-Score** gibt an, wie viele Standardabweichungen ein Wert vom Mittelwert entfernt ist.

---

1) Was ist eine Normalverteilung?

Daten können auf verschiedene Arten verteilt sein – sie können nach links oder rechts verzerrt sein, oder chaotisch durcheinander liegen. In vielen realen Situationen tendieren Messwerte jedoch dazu, sich um einen **zentralen Wert** zu gruppieren, ohne eine Tendenz nach links oder rechts zu zeigen. Diese Form nennt man **Normalverteilung**.

Sie sieht aus wie eine **Glocke** und wird deshalb oft auch **Glockenkurve** genannt.

1.1 Eigenschaften der Normalverteilung

Eine Normalverteilung hat folgende Eigenschaften:

- **Mittelwert = Median = Modus** – alle drei Lagenmasse fallen auf denselben Punkt.
- Die Kurve ist **symmetrisch** um den Mittelpunkt.
- **50 %** der Werte liegen unterhalb des Mittelwerts, **50 %** darüber.

1.2 Beispiele aus der Praxis

Viele reale Phänomene folgen näherungsweise einer Normalverteilung:

- Körpergrössen von Menschen
- Abmessungen von Industrieprodukten
- Messfehler
- Blutdruck
- Testergebnisse

> **Wichtig:** Die meisten Daten sind nicht *perfekt* normalverteilt, aber die Normalverteilung hilft uns trotzdem, sie zu verstehen und Aussagen darüber zu machen.

---

2) Standardabweichungen und die 68-95-99.7-Regel

Die **Standardabweichung** σ (Sigma) ist ein Mass dafür, wie stark die Werte um den Mittelwert streuen.

Bei einer Normalverteilung gilt näherungsweise:

Bereich	Anteil der Werte
Mittelwert ± 1 Standardabweichung	**68 %**
Mittelwert ± 2 Standardabweichungen	**95 %**
Mittelwert ± 3 Standardabweichungen	**99.7 %**

Diese Regel ist so wichtig, dass sie einen eigenen Namen hat: die **68-95-99.7-Regel** (oder Empirische Regel).

Beispiel: Körpergrössen in einer Schulklasse

In einer Schule sind **95 % der Schülerinnen und Schüler zwischen 1.1 m und 1.7 m gross**.

Angenommen, die Daten sind normalverteilt – wie gross sind Mittelwert und Standardabweichung?

Mittelwert: Der Mittelwert liegt in der Mitte:

$$ \mu = \frac{1.1 + 1.7}{2} = 1.4 \text{ m} $$

Standardabweichung: 95 % entsprechen ± 2 Standardabweichungen (also insgesamt 4 Standardabweichungen):

$$ \sigma = \frac{1.7 - 1.1}{4} = \frac{0.6}{4} = 0.15 \text{ m} $$

Ein Wert von 1.7 m liegt also genau **2 Standardabweichungen** über dem Mittelwert von 1.4 m.

> **Faustregel:**
>
> Ein Wert innerhalb von 1σ ist **wahrscheinlich** (68 von 100).
>
> Ein Wert innerhalb von 2σ ist **sehr wahrscheinlich** (95 von 100).
>
> Ein Wert innerhalb von 3σ ist **fast sicher** (997 von 1000).

---

3) Standardisierung und der z-Score

Die Anzahl Standardabweichungen, die ein Wert vom Mittelwert entfernt ist, heisst **z-Score** (oder Standardwert).

$$ z = \frac{x - \mu}{\sigma} $$

Dabei gilt:

- z = z-Score (Standardwert)
- x = der zu standardisierende Wert
- μ (Mu) = Mittelwert der Verteilung
- σ (Sigma) = Standardabweichung

3.1 Vorgehen: Wert standardisieren

1. Mittelwert subtrahieren: $x - \mu$
2. Durch die Standardabweichung dividieren: $\div \; \sigma$

3.2 Beispiel: Reisezeit

Eine Umfrage über tägliche Reisezeiten (in Minuten) ergab:

```
26, 33, 65, 28, 34, 55, 25, 44, 50, 36, 26, 37, 43, 62, 35, 38, 45, 32, 28, 34
```

Mittelwert: $\mu = 38.8$ Minuten
Standardabweichung: $\sigma = 11.4$ Minuten

Die ersten drei Werte werden wie folgt standardisiert:

Originalwert x	Berechnung	z-Score
26	$\frac{26 - 38.8}{11.4}$	**−1.12**
33	$\frac{33 - 38.8}{11.4}$	**−0.51**
65	$\frac{65 - 38.8}{11.4}$	**+2.30**

Ein z-Score von −1.12 bedeutet: Der Wert 26 liegt **1.12 Standardabweichungen unterhalb** des Mittelwerts.

3.3 Wozu dient die Standardisierung?

Durch die Standardisierung können wir **beliebige Normalverteilungen** in die **Standardnormalverteilung** umrechnen (Mittelwert = 0, Standardabweichung = 1). Das erlaubt uns:

- Werte aus verschiedenen Verteilungen direkt zu **vergleichen**.
- Mit einer einzigen **Standardnormaltabelle** Wahrscheinlichkeiten abzulesen, anstatt für jede Verteilung neu zu rechnen.

3.4 Beispiel: Testergebnisse fair beurteilen

Ein Lehrer bewertet einen Test (maximale Punktzahl: 60):

```
20, 15, 26, 32, 18, 28, 35, 14, 26, 22, 17
```

Die meisten Schülerinnen und Schüler haben weniger als 30 Punkte – der Test war offensichtlich sehr schwer. Der Lehrer entscheidet, die Noten zu standardisieren und nur Personen mit einem z-Score unter −1 als ungenügend zu bewerten.

- Mittelwert: \(\mu = 23\)
- Standardabweichung: \(\sigma = 6.6\)
- z-Scores: −0.45, **−1.21**, 0.45, 1.36, −0.76, 0.76, 1.82, **−1.36**, 0.45, −0.15, −0.91

Resultat: Nur **2 Schülerinnen/Schüler** fallen unter −1 und gelten als ungenügend – deutlich fairer!

---

4) Verteilung in Prozentzonen

Die Standardnormalverteilung lässt sich in Prozentzonen aufteilen, jeweils in Schritten von 0.5 Standardabweichungen:

Zone	Anteil in diesem Bereich	Kumulierter Anteil (von links)
unter −3σ	0.1 %	0.1 %
−3σ bis −2.5σ	0.5 %	0.6 %
−2σ bis −1.5σ	6.1 %	ca. 6.7 %
−1σ bis −0.5σ	15.0 %	ca. 30.9 %
−0.5σ bis 0	19.1 %	ca. 50.0 %
0 bis +0.5σ	19.1 %	ca. 69.1 %
+0.5σ bis +1σ	15.0 %	ca. 84.1 %
+1σ bis +1.5σ	9.2 %	ca. 93.3 %
+2σ bis +2.5σ	3.8 %	ca. 98.8 %
über +3σ	0.1 %	99.9 %

Beispiel: Wie gut war dein Testergebnis?

Dein Ergebnis liegt **0.5 Standardabweichungen über dem Durchschnitt** (z = +0.5). Wie viele Personen haben schlechter abgeschnitten?

- Anteil zwischen 0 und +0.5: **19.1 %**
- Anteil unter 0 (linke Hälfte): **50.0 %**
- Total: **69.1 %** haben schlechter abgeschnitten als du.

---

5) Praxisbeispiel: Zuckerbeutel abfüllen

Eine Maschine füllt Zuckerbeutel ab. Eine Stichprobe ergibt:

- Mittelwert: $\mu = 1010$ g
- Standardabweichung: $\sigma = 20$ g

Problem: Etwa 31 % der Beutel enthalten weniger als 1000 g – das ist zu viel!

Lösung: Wir wollen, dass 1000 g bei −2.5 Standardabweichungen liegt (nur 0.6 % zu wenig).

Option A – Mittelwert erhöhen:

$$ \text{Neuer Mittelwert} = 1000 + 2.5 \times 20 = 1000 + 50 = 1050 \text{ g} $$

Option B – Streuung verringern:

$$ \sigma_{neu} = \frac{10}{2.5} = 4 \text{ g} $$

Bei einer Standardabweichung von nur 4 g würden weniger als 0.6 % der Beutel das Gewichtslimit unterschreiten.

> **Fazit:** Mit der Normalverteilung können wir Produktionsprozesse gezielt steuern – entweder durch Anpassen des Mittelwerts oder durch Verringern der Streuung.

Verständnisfragen

1. Nenne **drei reale Phänomene**, die näherungsweise normalverteilt sind.
2. Ein Datensatz hat $\mu = 50$ und $\sigma = 10$. Welcher Anteil der Werte liegt zwischen 30 und 70?
3. Berechne den z-Score für einen Wert von $x = 75$, wenn $\mu = 60$ und $\sigma = 8$.
4. Was bedeutet ein z-Score von **−2.5** anschaulich?
5. Warum ist die Standardisierung nützlich, wenn man Werte aus verschiedenen Verteilungen vergleichen will?
